

Impact des algorithmes en IA dits « éthiques by design »

Jean-Marie John-Mathews

Christine Balagué, Dominique Cardon

IMT-BS, Sciences Po

1. CONTEXT

Les algorithmes constituent des intermédiaires de plus en plus présents dans les interactions entre les entreprises et les consommateurs mais aussi entre les individus eux-mêmes. Ce phénomène est source d'inquiétudes et de débats, ce qui a donné lieu ces dernières années à la production de plusieurs rapports et missions gouvernementaux et au développement dans plusieurs pays d'un courant de recherche interdisciplinaire sur la nécessité de développer des algorithmes éthiques by design.

2. CHALLENGES or OBJECTIVES or GAPS or NEED or RESEARCH QUESTIONS

Cette thèse aura pour premier but de clarifier les concepts autour de l'éthique des algorithmes en les inscrivant dans leur contexte d'utilisation (entreprises/consommateurs) et leur contexte d'élaboration technique. Le second but de ce travail de recherche est d'évaluer l'impact des méthodes de conception éthique by design des algorithmes à la fois sur les consommateurs finaux et sur les entreprises conceptrices de ces technologies. Des enjeux sur les outils techniques de certification de critères éthiques seront explorés, à l'état de recherche encore actuellement. Cette évaluation des impacts nous appellera à penser un modèle de gouvernance pour les entreprises ainsi qu'un modèle de régulation pour les pouvoirs publics afin de répondre aux différents enjeux éthiques.

3. CONTRIBUTION or PROPOSITION or RESEARCH METHOD

- Quelle est la cartographie des grandes préoccupations éthiques autour des algorithmes ?
- Comment paramétrise-t-on l'éthique des algorithmes ?
- Comment peut-on classifier ces paramétrisations ?
- Comment les utilisateurs veulent-ils être calculés ?

4. CASE STUDY or APPLICATION (or EXPECTED)

- Expérimentation auprès d'élèves ingénieurs afin d'étudier leur paramétrisation éthique
- Récupération des compromis éthiques auprès des utilisateurs
- Moral machine orientés algorithmes responsables

5. RESULTS or EXPECTED RESULTS

- L'usage des algorithmes d'IA n'est pas si répandu (à part les algorithmes des GAFAs) pour espérer une expérimentation sur utilisateur
- Les utilisateurs sur-déclarent leurs préférences éthiques dans le domaine du numérique

6. FUTURE WORK

- Finalisation des cartographies
- Mise en place des expérimentations

REFERENCES

- Doshi-Velez, and Kim. 2017. "Towards A Rigorous Science of Interpretable Machine Learning." *arXiv :1702.08608*.
- Dwork, Cynthia. 2012. "Fairness Through Awareness." *Proc. 3rd Innovations Theoretical Computer Sci. Conf.*
- Kroll, Joshua, Johanna Huey, Solon Barocas, Edward W. Felten, Joel R. Reidenberg, David G. Robinson, and Harlan Yu. 2017. "Accountable Algorithms." *University of Pennsylvania Law Review* 633-755.
- Mittelstadt, Brendt Daniel, Patrick Allo, Mariarosaria Taddeo, Sandra Wachter, and Luciano Floridi. 2016. "The ethics of algorithms: Mapping the debate." *Big Data & Society* 1-21.
- Zemel, Richard. 2013. "Learning Fair Representations." <http://proceedings.mlr.press/v28/zemel13.pdf>.